

Configuration	Epochs	Batch	Loss	Memory Usage (MB)	Elapsed Time (s)
learning_rate: 0.01 beta_1: 0.9 beta_2: 0.999 amsgrad: False	16	256	3833399552.0	3.754429817199707	4.39609
learning_rate: 0.01 beta_1: 0.9 beta_2: 0.999 amsgrad: True	16	256	3806195968.0	4.098898887634277	4.71422
learning_rate: 0.01 beta_1: 0.9 beta_2: 0.999999 amsgrad: False	16	256	3823437568.0	3.755620002746582	4.23564
learning_rate: 0.01 beta_1: 0.9 beta_2: 0.999999 amsgrad: True	16	256	3844052992.0	4.096096038818359	5.04987
learning_rate: 0.01 beta_1: 0.9 beta_2: 0.95 amsgrad: False	16	256	3792787968.0	3.7555551528930664	4.42226
learning_rate: 0.01 beta_1: 0.9 beta_2: 0.95 amsgrad: True	16	256	3856265216.0	4.095673561096191	5.51689
learning_rate: 0.01 beta_1: 0.9 beta_2: 0.9 amsgrad: False	16	256	3809671424.0	3.7569189071655273	5.56442
learning_rate: 0.01 beta_1: 0.9 beta_2: 0.9 amsgrad: True	16	256	3856465920.0	4.100458145141602	4.64079
learning_rate: 0.01 beta_1: 0.85 beta_2: 0.999 amsgrad: False	16	256	3842789376.0	3.7595720291137695	4.35549
learning_rate: 0.01 beta_1: 0.85 beta_2: 0.999 amsgrad: True	16	256	3910114048.0	4.093563079833984	4.52763

learning_rate: 0.01 beta_1: 0.85 beta_2: 0.999999 amsgrad: False	16	256	3854982144.0	3.7566394805908203	4.38271
learning_rate: 0.01 beta_1: 0.85 beta_2: 0.999999 amsgrad: True	16	256	3866231040.0	4.092703819274902	4.66346
learning_rate: 0.01 beta_1: 0.85 beta_2: 0.95 amsgrad: False	16	256	3810001152.0	3.7554140090942383	4.86787
learning_rate: 0.01 beta_1: 0.85 beta_2: 0.95 amsgrad: True	16	256	3889040128.0	4.099913597106934	4.63842
learning_rate: 0.01 beta_1: 0.85 beta_2: 0.9 amsgrad: False	16	256	3821802496.0	3.7555418014526367	5.69954
learning_rate: 0.01 beta_1: 0.85 beta_2: 0.9 amsgrad: True	16	256	3875096320.0	4.096251487731934	4.91814
learning_rate: 0.01 beta_1: 0.8 beta_2: 0.999 amsgrad: False	16	256	3926415872.0	3.7558937072753906	4.33837
learning_rate: 0.01 beta_1: 0.8 beta_2: 0.999 amsgrad: True	16	256	3821961728.0	4.094403266906738	4.88177
learning_rate: 0.01 beta_1: 0.8 beta_2: 0.999999 amsgrad: False	16	256	3918337536.0	3.758185386657715	4.67253
learning_rate: 0.01 beta_1: 0.8 beta_2: 0.999999 amsgrad: True	16	256	3801870080.0	4.094328880310059	4.55311

learning_rate: 0.01 beta_1: 0.8 beta_2: 0.95 amsgrad: False	16	256	3800365312.0	3.7596025466918945	4.68644
learning_rate: 0.01 beta_1: 0.8 beta_2: 0.95 amsgrad: True	16	256	3870151168.0	4.095641136169434	4.65346
learning_rate: 0.01 beta_1: 0.8 beta_2: 0.9 amsgrad: False	16	256	3873157376.0	3.756150245666504	4.21799
learning_rate: 0.01 beta_1: 0.8 beta_2: 0.9 amsgrad: True	16	256	3850761728.0	4.094635009765625	4.80574
learning_rate: 0.01 beta_1: 0.75 beta_2: 0.999 amsgrad: False	16	256	3846341120.0	3.756680488586426	4.52351
learning_rate: 0.01 beta_1: 0.75 beta_2: 0.999 amsgrad: True	16	256	4230362112.0	4.098578453063965	4.58045
learning_rate: 0.01 beta_1: 0.75 beta_2: 0.999999 amsgrad: False	16	256	3877543936.0	3.7561302185058594	4.32702
learning_rate: 0.01 beta_1: 0.75 beta_2: 0.999999 amsgrad: True	16	256	3833294848.0	4.095860481262207	5.19713
learning_rate: 0.01 beta_1: 0.75 beta_2: 0.95 amsgrad: False	16	256	3930062592.0	3.75677490234375	4.30622
learning_rate: 0.01 beta_1: 0.75 beta_2: 0.95 amsgrad: True	16	256	3881378816.0	4.099329948425293	7.84701

learning_rate: 0.01 beta_1: 0.75 beta_2: 0.9 amsgrad: False	16	256	3832943616.0	3.754897117614746	4.70773
learning_rate: 0.01 beta_1: 0.75 beta_2: 0.9 amsgrad: True	16	256	3889588480.0	4.094521522521973	5.2096
learning_rate: 0.01 beta_1: 0.5 beta_2: 0.999 amsgrad: False	16	256	3780736000.0	3.756850242614746	4.22083
learning_rate: 0.01 beta_1: 0.5 beta_2: 0.999 amsgrad: True	16	256	3805395200.0	4.094820022583008	4.53678
learning_rate: 0.01 beta_1: 0.5 beta_2: 0.999999 amsgrad: False	16	256	3854842880.0	3.7543153762817383	4.21615
learning_rate: 0.01 beta_1: 0.5 beta_2: 0.999999 amsgrad: True	16	256	3784662528.0	4.096040725708008	4.5605
learning_rate: 0.01 beta_1: 0.5 beta_2: 0.95 amsgrad: False	16	256	3850976000.0	3.75698184967041	4.19463
learning_rate: 0.01 beta_1: 0.5 beta_2: 0.95 amsgrad: True	16	256	3840916736.0	4.094295501708984	4.49459
learning_rate: 0.01 beta_1: 0.5 beta_2: 0.9 amsgrad: False	16	256	3860883968.0	3.7566490173339844	4.21091
learning_rate: 0.01 beta_1: 0.5 beta_2: 0.9 amsgrad: True	16	256	3963157248.0	4.09639835357666	4.68269

learning_rate: 0.001 beta_1: 0.9 beta_2: 0.999 amsgrad: False	16	256	8723546112.0	3.758066177368164	4.1712
learning_rate: 0.001 beta_1: 0.9 beta_2: 0.999 amsgrad: True	16	256	8653972480.0	4.0949811935424805	4.50148
learning_rate: 0.001 beta_1: 0.9 beta_2: 0.999999 amsgrad: False	16	256	8045557248.0	3.756585121154785	4.14015
learning_rate: 0.001 beta_1: 0.9 beta_2: 0.999999 amsgrad: True	16	256	8397444608.0	4.095656394958496	4.70863
learning_rate: 0.001 beta_1: 0.9 beta_2: 0.95 amsgrad: False	16	256	12509367296.0	3.7547359466552734	4.2457
learning_rate: 0.001 beta_1: 0.9 beta_2: 0.95 amsgrad: True	16	256	14626565120.0	4.0942535400390625	4.471
learning_rate: 0.001 beta_1: 0.9 beta_2: 0.9 amsgrad: False	16	256	12303429632.0	3.7559938430786133	4.20523
learning_rate: 0.001 beta_1: 0.9 beta_2: 0.9 amsgrad: True	16	256	15427702784.0	4.096846580505371	4.55935
learning_rate: 0.001 beta_1: 0.85 beta_2: 0.999 amsgrad: False	16	256	8352675840.0	3.7576208114624023	4.18954
learning_rate: 0.001 beta_1: 0.85 beta_2: 0.999 amsgrad: True	16	256	8461594624.0	4.093345642089844	4.48217

learning_rate: 0.001 beta_1: 0.85 beta_2: 0.999999 amsgrad: False	16	256	8056255488.0	3.7547359466552734	4.18901
learning_rate: 0.001 beta_1: 0.85 beta_2: 0.999999 amsgrad: True	16	256	8163605504.0	4.09581184387207	4.58281
learning_rate: 0.001 beta_1: 0.85 beta_2: 0.95 amsgrad: False	16	256	11984776192.0	3.756795883178711	4.17825
learning_rate: 0.001 beta_1: 0.85 beta_2: 0.95 amsgrad: True	16	256	13318728704.0	4.094799041748047	4.46822
learning_rate: 0.001 beta_1: 0.85 beta_2: 0.9 amsgrad: False	16	256	11876264960.0	3.757854461669922	4.14115
learning_rate: 0.001 beta_1: 0.85 beta_2: 0.9 amsgrad: True	16	256	15780306944.0	4.096987724304199	4.61162
learning_rate: 0.001 beta_1: 0.8 beta_2: 0.999 amsgrad: False	16	256	8177865216.0	3.7549495697021484	4.23358
learning_rate: 0.001 beta_1: 0.8 beta_2: 0.999 amsgrad: True	16	256	8610812928.0	4.097339630126953	4.56748
learning_rate: 0.001 beta_1: 0.8 beta_2: 0.999999 amsgrad: False	16	256	8308609536.0	3.7559728622436523	4.18495
learning_rate: 0.001 beta_1: 0.8 beta_2: 0.999999 amsgrad: True	16	256	8220268032.0	4.094504356384277	4.63709

learning_rate: 0.001 beta_1: 0.8 beta_2: 0.95 amsgrad: False	16	256	12077587456.0	3.757685661315918	4.27819
learning_rate: 0.001 beta_1: 0.8 beta_2: 0.95 amsgrad: True	16	256	12608310272.0	4.093216896057129	5.44491
learning_rate: 0.001 beta_1: 0.8 beta_2: 0.9 amsgrad: False	16	256	12028633088.0	3.7546377182006836	4.28425
learning_rate: 0.001 beta_1: 0.8 beta_2: 0.9 amsgrad: True	16	256	14783427584.0	4.100025177001953	4.82082
learning_rate: 0.001 beta_1: 0.75 beta_2: 0.999 amsgrad: False	16	256	8655734784.0	3.756241798400879	4.58951
learning_rate: 0.001 beta_1: 0.75 beta_2: 0.999 amsgrad: True	16	256	8740887552.0	4.098150253295898	4.65386
learning_rate: 0.001 beta_1: 0.75 beta_2: 0.999999 amsgrad: False	16	256	8349539840.0	3.7570133209228516	4.4215
learning_rate: 0.001 beta_1: 0.75 beta_2: 0.999999 amsgrad: True	16	256	8480273920.0	4.098308563232422	5.07704
learning_rate: 0.001 beta_1: 0.75 beta_2: 0.95 amsgrad: False	16	256	11196273664.0	3.7584991455078125	4.25314
learning_rate: 0.001 beta_1: 0.75 beta_2: 0.95 amsgrad: True	16	256	12908059648.0	4.092887878417969	4.74307

learning_rate: 0.001 beta_1: 0.75 beta_2: 0.9 amsgrad: False	16	256	13298925568.0	3.756442070007324	4.21036
learning_rate: 0.001 beta_1: 0.75 beta_2: 0.9 amsgrad: True	16	256	15036506112.0	4.0975446701049805	4.97253
learning_rate: 0.001 beta_1: 0.5 beta_2: 0.999 amsgrad: False	16	256	8777032704.0	3.753826141357422	4.51335
learning_rate: 0.001 beta_1: 0.5 beta_2: 0.999 amsgrad: True	16	256	8831939584.0	4.093005180358887	4.47877
learning_rate: 0.001 beta_1: 0.5 beta_2: 0.999999 amsgrad: False	16	256	8362523648.0	3.7561845779418945	4.1601
learning_rate: 0.001 beta_1: 0.5 beta_2: 0.999999 amsgrad: True	16	256	8310107136.0	4.094951629638672	4.54598
learning_rate: 0.001 beta_1: 0.5 beta_2: 0.95 amsgrad: False	16	256	11502692352.0	3.7586545944213867	4.19968
learning_rate: 0.001 beta_1: 0.5 beta_2: 0.95 amsgrad: True	16	256	14166900736.0	4.0994977951049805	4.5223
learning_rate: 0.001 beta_1: 0.5 beta_2: 0.9 amsgrad: False	16	256	12464422912.0	3.757706642150879	4.23328
learning_rate: 0.001 beta_1: 0.5 beta_2: 0.9 amsgrad: True	16	256	14790241280.0	4.0948028564453125	4.51895

learning_rate: 0.0001 beta_1: 0.9 beta_2: 0.999 amsgrad: False	16	256	51138121728.0	3.756378173828125	4.3319
learning_rate: 0.0001 beta_1: 0.9 beta_2: 0.999 amsgrad: True	16	256	51099553792.0	4.095726013183594	4.59393
learning_rate: 0.0001 beta_1: 0.9 beta_2: 0.999999 amsgrad: False	16	256	50807644160.0	3.756129264831543	4.22763
learning_rate: 0.0001 beta_1: 0.9 beta_2: 0.999999 amsgrad: True	16	256	50830991360.0	4.093208312988281	4.49406
learning_rate: 0.0001 beta_1: 0.9 beta_2: 0.95 amsgrad: False	16	256	51486175232.0	3.7569456100463867	4.21809
learning_rate: 0.0001 beta_1: 0.9 beta_2: 0.95 amsgrad: True	16	256	51478007808.0	4.09975528717041	4.63015
learning_rate: 0.0001 beta_1: 0.9 beta_2: 0.9 amsgrad: False	16	256	51502125056.0	3.754572868347168	4.20863
learning_rate: 0.0001 beta_1: 0.9 beta_2: 0.9 amsgrad: True	16	256	51513225216.0	4.0978546142578125	4.58453
learning_rate: 0.0001 beta_1: 0.85 beta_2: 0.999 amsgrad: False	16	256	51008536576.0	3.755709648132324	4.22947
learning_rate: 0.0001 beta_1: 0.85 beta_2: 0.999 amsgrad: True	16	256	51013009408.0	4.095044136047363	4.83236

learning_rate: 0.0001 beta_1: 0.85 beta_2: 0.999999 amsgrad: False	16	256	50785812480.0	3.756563186645508	4.24541
learning_rate: 0.0001 beta_1: 0.85 beta_2: 0.999999 amsgrad: True	16	256	50842181632.0	4.095701217651367	4.58203
learning_rate: 0.0001 beta_1: 0.85 beta_2: 0.95 amsgrad: False	16	256	51473203200.0	3.7551393508911133	4.19064
learning_rate: 0.0001 beta_1: 0.85 beta_2: 0.95 amsgrad: True	16	256	51492749312.0	4.094491004943848	4.64925
learning_rate: 0.0001 beta_1: 0.85 beta_2: 0.9 amsgrad: False	16	256	51478085632.0	3.755051612854004	4.24745
learning_rate: 0.0001 beta_1: 0.85 beta_2: 0.9 amsgrad: True	16	256	51522629632.0	4.09881591796875	4.49944
learning_rate: 0.0001 beta_1: 0.8 beta_2: 0.999 amsgrad: False	16	256	50989969408.0	3.7571840286254883	4.23798
learning_rate: 0.0001 beta_1: 0.8 beta_2: 0.999 amsgrad: True	16	256	51025211392.0	4.0940961837768555	4.52658
learning_rate: 0.0001 beta_1: 0.8 beta_2: 0.999999 amsgrad: False	16	256	50849796096.0	3.756723403930664	4.26406
learning_rate: 0.0001 beta_1: 0.8 beta_2: 0.999999 amsgrad: True	16	256	50903568384.0	4.095769882202148	4.50647

learning_rate: 0.0001 beta_1: 0.8 beta_2: 0.95 amsgrad: False	16	256	51471929344.0	3.7556686401367188	4.1965
learning_rate: 0.0001 beta_1: 0.8 beta_2: 0.95 amsgrad: True	16	256	51491098624.0	4.092007637023926	4.49968
learning_rate: 0.0001 beta_1: 0.8 beta_2: 0.9 amsgrad: False	16	256	51502211072.0	3.755526542663574	4.29289
learning_rate: 0.0001 beta_1: 0.8 beta_2: 0.9 amsgrad: True	16	256	51509542912.0	4.0980119705200195	4.75872
learning_rate: 0.0001 beta_1: 0.75 beta_2: 0.999 amsgrad: False	16	256	51128705024.0	3.757500648498535	4.28768
learning_rate: 0.0001 beta_1: 0.75 beta_2: 0.999 amsgrad: True	16	256	51067482112.0	4.094099044799805	4.6016
learning_rate: 0.0001 beta_1: 0.75 beta_2: 0.999999 amsgrad: False	16	256	50919489536.0	3.7544946670532227	4.19828
learning_rate: 0.0001 beta_1: 0.75 beta_2: 0.999999 amsgrad: True	16	256	50880151552.0	4.094250679016113	4.66263
learning_rate: 0.0001 beta_1: 0.75 beta_2: 0.95 amsgrad: False	16	256	51469946880.0	3.754857063293457	4.56552
learning_rate: 0.0001 beta_1: 0.75 beta_2: 0.95 amsgrad: True	16	256	51505905664.0	4.097536087036133	4.5213

learning_rate: 0.0001 beta_1: 0.75 beta_2: 0.9 amsgrad: False	16	256	51497361408.0	3.7552804946899414	4.27245
learning_rate: 0.0001 beta_1: 0.75 beta_2: 0.9 amsgrad: True	16	256	51532312576.0	4.095497131347656	4.48286
learning_rate: 0.0001 beta_1: 0.5 beta_2: 0.999 amsgrad: False	16	256	51029385216.0	3.7559690475463867	4.25847
learning_rate: 0.0001 beta_1: 0.5 beta_2: 0.999 amsgrad: True	16	256	51048194048.0	4.093844413757324	4.53115
learning_rate: 0.0001 beta_1: 0.5 beta_2: 0.999999 amsgrad: False	16	256	50776268800.0	3.7575817108154297	4.19462
learning_rate: 0.0001 beta_1: 0.5 beta_2: 0.999999 amsgrad: True	16	256	50877169664.0	4.100652694702148	4.55671
learning_rate: 0.0001 beta_1: 0.5 beta_2: 0.95 amsgrad: False	16	256	51454812160.0	3.7554759979248047	4.22353
learning_rate: 0.0001 beta_1: 0.5 beta_2: 0.95 amsgrad: True	16	256	51495460864.0	4.0948638916015625	4.52724
learning_rate: 0.0001 beta_1: 0.5 beta_2: 0.9 amsgrad: False	16	256	51458826240.0	3.756040573120117	4.55066
learning_rate: 0.0001 beta_1: 0.5 beta_2: 0.9 amsgrad: True	16	256	51512815616.0	4.099480628967285	4.48689

learning_rate: 1e-05 beta_1: 0.9 beta_2: 0.999 amsgrad: False	16	256	51739967488.0	3.7542171478271484	4.17627
learning_rate: 1e-05 beta_1: 0.9 beta_2: 0.999 amsgrad: True	16	256	51741282304.0	4.093377113342285	4.55587
learning_rate: 1e-05 beta_1: 0.9 beta_2: 0.999999 amsgrad: False	16	256	51740626944.0	3.7568225860595703	4.1742
learning_rate: 1e-05 beta_1: 0.9 beta_2: 0.999999 amsgrad: True	16	256	51741782016.0	4.09869384765625	4.5256
learning_rate: 1e-05 beta_1: 0.9 beta_2: 0.95 amsgrad: False	16	256	51743334400.0	3.754213333129883	4.53598
learning_rate: 1e-05 beta_1: 0.9 beta_2: 0.95 amsgrad: True	16	256	51744546816.0	4.093557357788086	4.49715
learning_rate: 1e-05 beta_1: 0.9 beta_2: 0.9 amsgrad: False	16	256	51743027200.0	3.756612777709961	4.23136
learning_rate: 1e-05 beta_1: 0.9 beta_2: 0.9 amsgrad: True	16	256	51744423936.0	4.0955705642700195	4.56746
learning_rate: 1e-05 beta_1: 0.85 beta_2: 0.999 amsgrad: False	16	256	51742748672.0	3.7562170028686523	4.61419
learning_rate: 1e-05 beta_1: 0.85 beta_2: 0.999 amsgrad: True	16	256	51741003776.0	4.095767021179199	4.55808

learning_rate: 1e-05 beta_1: 0.85 beta_2: 0.999999 amsgrad: False	16	256	51739451392.0	3.75673770904541	4.25772
learning_rate: 1e-05 beta_1: 0.85 beta_2: 0.999999 amsgrad: True	16	256	51739430912.0	4.094341278076172	4.53024
learning_rate: 1e-05 beta_1: 0.85 beta_2: 0.95 amsgrad: False	16	256	51742257152.0	3.7554588317871094	4.3958
learning_rate: 1e-05 beta_1: 0.85 beta_2: 0.95 amsgrad: True	16	256	51743715328.0	4.09294319152832	4.55554
learning_rate: 1e-05 beta_1: 0.85 beta_2: 0.9 amsgrad: False	16	256	51743682560.0	3.754404067993164	4.67548
learning_rate: 1e-05 beta_1: 0.85 beta_2: 0.9 amsgrad: True	16	256	51745824768.0	4.094164848327637	4.49439
learning_rate: 1e-05 beta_1: 0.8 beta_2: 0.999 amsgrad: False	16	256	51739893760.0	3.75862979888916	4.27419
learning_rate: 1e-05 beta_1: 0.8 beta_2: 0.999 amsgrad: True	16	256	51741057024.0	4.09330940246582	4.8334
learning_rate: 1e-05 beta_1: 0.8 beta_2: 0.999999 amsgrad: False	16	256	51740479488.0	3.7570037841796875	4.29676
learning_rate: 1e-05 beta_1: 0.8 beta_2: 0.999999 amsgrad: True	16	256	51738636288.0	4.099384307861328	4.64626

learning_rate: 1e-05 beta_1: 0.8 beta_2: 0.95 amsgrad: False	16	256	51743588352.0	3.757556915283203	4.3167
learning_rate: 1e-05 beta_1: 0.8 beta_2: 0.95 amsgrad: True	16	256	51743584256.0	4.096229553222656	4.66424
learning_rate: 1e-05 beta_1: 0.8 beta_2: 0.9 amsgrad: False	16	256	51744739328.0	3.7550697326660156	4.19488
learning_rate: 1e-05 beta_1: 0.8 beta_2: 0.9 amsgrad: True	16	256	51744653312.0	4.095026016235352	4.84795
learning_rate: 1e-05 beta_1: 0.75 beta_2: 0.999 amsgrad: False	16	256	51741896704.0	3.7546281814575195	4.21732
learning_rate: 1e-05 beta_1: 0.75 beta_2: 0.999 amsgrad: True	16	256	51741114368.0	4.0991106033325195	4.5317
learning_rate: 1e-05 beta_1: 0.75 beta_2: 0.999999 amsgrad: False	16	256	51740983296.0	3.755809783935547	4.22633
learning_rate: 1e-05 beta_1: 0.75 beta_2: 0.999999 amsgrad: True	16	256	51740979200.0	4.093195915222168	4.52135
learning_rate: 1e-05 beta_1: 0.75 beta_2: 0.95 amsgrad: False	16	256	51743645696.0	3.75634765625	4.2035
learning_rate: 1e-05 beta_1: 0.75 beta_2: 0.95 amsgrad: True	16	256	51743666176.0	4.092268943786621	4.82771

learning_rate: 1e-05 beta_1: 0.75 beta_2: 0.9 amsgrad: False	16	256	51742932992.0	3.7550039291381836	4.21566
learning_rate: 1e-05 beta_1: 0.75 beta_2: 0.9 amsgrad: True	16	256	51743920128.0	4.095605850219727	4.52278
learning_rate: 1e-05 beta_1: 0.5 beta_2: 0.999 amsgrad: False	16	256	51742060544.0	3.757462501525879	4.33366
learning_rate: 1e-05 beta_1: 0.5 beta_2: 0.999 amsgrad: True	16	256	51740954624.0	4.098522186279297	4.58641
learning_rate: 1e-05 beta_1: 0.5 beta_2: 0.999999 amsgrad: False	16	256	51740712960.0	3.757486343383789	4.22139
learning_rate: 1e-05 beta_1: 0.5 beta_2: 0.999999 amsgrad: True	16	256	51740925952.0	4.093489646911621	4.71271
learning_rate: 1e-05 beta_1: 0.5 beta_2: 0.95 amsgrad: False	16	256	51743379456.0	3.756134033203125	4.28289
learning_rate: 1e-05 beta_1: 0.5 beta_2: 0.95 amsgrad: True	16	256	51744772096.0	4.100917816162109	4.54006
learning_rate: 1e-05 beta_1: 0.5 beta_2: 0.9 amsgrad: False	16	256	51743817728.0	3.755377769470215	4.20769
learning_rate: 1e-05 beta_1: 0.5 beta_2: 0.9 amsgrad: True	16	256	51745128448.0	4.095083236694336	4.56376